PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06F 17/50

A1

(11) International Publication Number: WO 95/06293

(43) International Publication Date: 2 March 1995 (02.03.95)

(21) International Application Number: PCT/IB94/00257

(22) International Filing Date: 25 August 1994 (25.08.94)

(30) Priority Data:
0960/93 25 August 1993 (25.08.93) DK

(71) Applicant *(for all designated States except US)*: SYMBICOM AKTIEBOLAG [SE/SE]; Tvistevägen 48, P.O. Box 1451, S-901 24 Umeå (SE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: MEDINA, Carmen [MX/SE]; Flogstavägen 33B, S-752 64 Uppsala (SE). AQVIST, Johan [SE/SE]; Grågåsvägen 2, S-756 51 Uppsala (SE).

(74) Agent: PLOUGMANN & VINGTOFT A/S; Sankt Annæ Plads 11, P.O. Box 3007, DK-1021 Copenhagen K (DK).

(81) Designated States: AM, AU, BB, BG, BR, BY, CA, CN, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, FI, FI (Utility model), GE, HU, JP, KE, KG, KP, KR, KZ, LK, LT, LV, MD, MG, MN, MW, NO, NZ, PL, RO, RU, SD, SI, SK, SK (Utility model), TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD).

Published
With international search report.

(54) Title: MOLECULAR MODELLING AND DRUG DESIGN

(57) Abstract

A novel method for the determination of the binding free energy between a host or receptor molecule and a ligand substance is disclosed. The method comprises the determination of adjusted combinations of the average energy difference between the contribution from a) polar and b) non-polar interactions between the substance in two states, one where the substance alone and another where the substance bound to the host or receptor molecule is surrounded by solvent. The contributions from the polar and non-polar interactions are preferably provided by molecular dynamics simulations derived from 3-dimensional representations of the receptor molecule and the substance. The method renders possible the computation of binding free energies without the use of Monte Carlo simulations/free energy perturbation techniques and thus requires less computational resources than known methods. Also disclosed are methods for selecting/providing substances capable of interacting with a host or receptor molecule.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgyzstan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

MOLECULAR MODELLING AND DRUG DESIGN

FIELD OF INVENTION

The present invention relates to a method for assessing the absolute free binding energy between a host or receptor molecule and a chemical substance interacting therewith, e.g. bound thereto, and to a method for assessing the relative free binding energy between a plurality of systems comprising a host or receptor molecule and a chemical substance interacting therewith.

BACKGROUND OF THE INVENTION

Computational chemistry and molecular modelling have become an essential part of the modern drug design process. This type of methodology is now being used as a complement to experimental biochemical studies and 3-dimensional structure determination by crystallographic or NMR methods. The most important applications of computational modelling in drug design comprise (i) methods for finding new "lead compounds", (ii) interactive computer graphics for modifying and manipulating the chemical and geometrical structure of inhibitors and (iii) subsequent energy and structure refinement using molecular mechanics (MM) or molecular dynamics (MD) calculations (For reviews, see Cohen et al., 1990, Dixon, 1992). Until now, however, molecular modelling methods can only provide a certain extent of qualitative information on the affinity of various compounds for at given target site. At best, some tentative binding candidates can be proposed and perhaps some compounds can be excluded which do not seem to match the target site well. This may be done by considering molecular shape and electrostatic properties, but the quantitative discrimination between different ligands in terms of actual binding constants must usually be left to the experimental work of organic chemists. The reason for this is that it is extremely difficult to theoretically predict relative or absolute binding constants, i.e., *free energies*, for all

but the very simplest cases. The only available approach to this type of problem today is the so-called Free Energy Perturbation (FEP) technique (for reviews, see: Beveridge and DiCapua, 1989, Jorgensen, 1989, Straatsma and McCammon, 1992) which in practice is limited to such "small perturbations" that it cannot realistically be used for molecules of the sizes which drug compounds represent.

If a typical case is considered where it is desired to determine the *relative* free energy of binding between two compounds, A and B, the problem is described by the thermodynamic cycle:

$$\begin{array}{ccc} & \xrightarrow{\Delta\Delta G^{w}_{sol}} & \\ A(w) & \rightarrow & B(w) \\ \Delta G_{bind}(A) \downarrow & & \downarrow \Delta G_{bind}(B) \\ A(p) & \rightarrow & B(p) \\ & \xrightarrow{\Delta\Delta G^{p}_{sol}} & \end{array}$$

where $\Delta\Delta G^{w}_{sol}$ and $\Delta\Delta G^{p}_{sol}$ denote the differences in solvation energy between A and B in water and in the (solvated) protein site, respectively, and the ΔG_{bind}'s are the corresponding binding energies; $\Delta G_{bind}(B) - \Delta G_{bind}(A) = \Delta\Delta G^{p}_{sol} - \Delta\Delta G^{w}_{sol}$. The same cycle can be used to determine the *absolute* binding constant of B if A denotes a nil particle. In that case $\Delta G_{bind}(A) \equiv 0$ and the binding energy is obtained as the difference between the absolute solvation energies for B in water and in the protein. With the FEP approach, the free energies associated with the two unphysical paths A(w) → B(w) and A(p) → B(p) are calculated, corresponding to a "mutation" of A into B (or the creation of B in the case where A is a nil particle). MD (or Monte Carlo) simulations are used to collect ensemble averages along the paths, which must be rather fine grained in order for the free energy to converge properly. If, e.g., two enzyme inhibitors are considered, the path connecting them will involve changes in the molecular charge distribution as well as the creation/annihilation of atoms. Especially the latter type of process converges slowly

and thus, if the A and B "states" are very different, it will be difficult to obtain convergent free energies even with present day computer resources (Pearlman and Kollman, 1991; Straatsma and McCammon, 1991; Mitchell and McCammon, 1991). It should also be noted that most of the computing time in FEP is actually spent on uninteresting configurations that correspond to a "mixture" of A and B. Moreover, if large conformational changes are involved, this always poses a major problem. Hence, the method only really works well when A and B are rather similar and, in fact, all applications related to drug design that have been reported fall within this category (e.g., Wong and McCammon, 1986; Bash et al., 1987; Brooks, 1989; Merz and Kollmann, 1989; McDonald and Brooks, 1991; Rao and Singh; 1991; Tropsha and Hermans, 1992). The largest "perturbation" to which the FEP method has been applied is a hydrogen to hexyl group transformation reported by Merz et al., 1991. Drug design in practice, however, often deals with relatively large inhibitors that differ considerably from each other so that neither relative nor absolute free energies can be obtained with the FEP/MD method within reasonable computing time.

Thus, there is an urgent demand for a method for assessing free energies of binding between host or receptor molecules and chemical substances of sizes relevant for drug compounds without the necessity of synthesizing the chemical substances.

BRIEF DISCLOSURE OF INVENTION

The present invention provides a method which makes it possible to assess free energies of binding between a "host or receptor" molecule and a chemical substances, such as a drug candidate, interacting therewith, without the necessity of synthesizing the chemical substances. Thereby, the time, efforts and costs involved in arriving at realistic drug candidates which it is justified to actually synthesize can be dramatically reduced, and the researchers involved in the

drug development will have a greater degree of freedom in their work during the early and intermediate phases of the development work.

In the present context, the term "a host or receptor molecule" should be understood in a broad sense as any molecule which can interact with a chemical substance, and the interaction of which with a chemical substance or a group or plurality of chemical substances, e.g. drug candidates, is to be studied. Thus, the host or receptor molecule may simply be any another chemical compound capable of interacting with the chemical substance, but most often, the host or receptor molecule will be a relatively large molecule, in other words a macromolecule such as a protein or an oligonucleotide, which is relatively large compared to the chemical substance; although the chemical substance interacting with the host or receptor molecule may, of course, in itself be a macromolecule. In many cases relevant in practice, the host or receptor molecule is a relatively large molecule of natural origin or prepared by recombinant DNA technique and having a particular biological function, e.g. as an enzyme, an antigen, an antibody, a biological receptor, etc., and the chemical substance is a synthetic substance of a structure known or believed to interact with or bind to the host or receptor molecule.

Contrary to the Free Energy Perturbation technique, the method of the invention does not involve "mutational paths", but rather determines the free energy of binding by an approximation, suitably a linear approximation which *only* involves an average interaction between the chemical substance and its surrounding. The interaction (or potential) energy between the chemical substance (in the following often referred to as the "drug") and its surrounding is divided into a polar (electrostatic) and a non-polar (hydrophobic) contribution, and the absolute free energy of binding is assessed as an adjusted combination of these two contributions.

Each of these contributions to the absolute free energy of binding is assessed as the *difference* between two distinct states, A and B, of interaction between the chemical substance and its surroundings which define the binding process, one state (A) being a state in which the chemical substance is surrounded by solvent, and the other state (B) being a state in which the chemical substance, interacting with (bound to) the host or receptor molecule, is surrounded by solvent.

Thus, the method of the invention for assessing the absolute free energy of binding between a host or receptor molecule and a chemical substance comprises:

1) assessing the average energy difference, $\langle \Delta V_{i-s}^{el} \rangle$, defined as $\langle V_{i-s}^{el} \rangle_B - \langle V_{i-s}^{el} \rangle_A$, between the contribution from polar interactions to the potential energy between the chemical substance (denoted i) and its surroundings (denoted s) in two states, one state (A) being where the chemical substance is surrounded by solvent, the other state (B) being where the chemical substance, bound to the host or receptor molecule, is surrounded by solvent,

2) assessing the average energy difference, $\langle \Delta V_{i-s}^{vdw} \rangle$, defined as $\langle V_{i-s}^{vdw} \rangle_B - \langle V_{i-s}^{vdw} \rangle_A$, between the contribution from non-polar interactions to the potential energy between the chemical substance (denoted i) and its surroundings (denoted s) in two states, one state (A) being where the chemical substance is surrounded by solvent, the other state (B) being where the chemical substance, bound to the host or receptor molecule, is surrounded by solvent, and

3) calculating the absolute binding free energy as an adjusted combination of the two above-mentioned average energy differences.

According to one embodiment of the present invention, the adjusted combination of the above-mentioned energy differen-

ces comprises about one half of the value of the polar binding energy difference between states B and A, ΔV^{el}_{i-s}, preferably one half of this value, in accordance with the linear response approximation, e.g. Marcus' theory of electron transfer reactions (Marcus, 1964), to which is added the non-polar (hydrophobic) contribution, $\langle \Delta V^{vdw}_{i-s} \rangle$, adjusted by means of an empirical parameter.

Thus, it has been found that the above-mentioned combination of the average energy differences can productively be assumed to contribute to the free energy of binding according to

$$\Delta G_{bind} = \tfrac{1}{2}\langle \Delta V^{el}_{i-s} \rangle + \alpha \langle \Delta V^{vdw}_{i-s} \rangle \qquad (1)$$

where the Δ's denote the difference between the two states, A and B, which define the binding process. That is, A represents the "drug" free in solution and B represents the case with the "drug" interacting with the solvated protein.

Thus, $\langle \Delta V^{el}_{i-s} \rangle = \langle V^{el}_{i-s} \rangle_B - \langle V^{el}_{i-s} \rangle_A$ where $\langle \rangle_A$, $\langle \rangle_B$ denote molecular dynamics averages calculated for the corresponding state. Equation 1 can thus be written

$$\Delta G_{bind} = \tfrac{1}{2}(\langle V^{el}_{i-s} \rangle_B - \langle V^{el}_{i-s} \rangle_A) + \alpha(\langle V^{vdw}_{i-s} \rangle_B - \langle V^{vdw}_{i-s} \rangle_A)$$

Each of the four averages can be calculated by standard molecular dynamics procedures using suitable computer soft- and hardware.

In the mathematical equations herein, the symbol < > means molecular dynamics average. The index i-s means compound-solvent (or compound-surrounding), the letter "i" standing for "inhibitor", which reflects the fact that the relevant compound or drug will often be a compound or drug which is intended to inhibit the function of the host or receptor molecule. The superscript "el" designates the polar or electrostatic energy, while the superscript "vdw" indicates "van der Waals", another designation for the non-polar interac-

tions. The symbol Δ indicates that the quantity in state A is subtracted from the quantity in state B.

Considerations in Connection with the Development of a Preferred Embodiment of the Invention

As a starting point was taken the linear response approximation for electrostatic forces which for polar solutions as a result yields quadratic free energy functions in response to the development of charges. This is, e.g., the familiar result from Marcus' theory of electron transfer reactions (Marcus, 1964). For a system with two states, A and B, given by two potential energy functions V_A and V_B one obtains, within the approximation of harmonic free energy functions of equal curvature, the relationship (see Lee et al., 1992 and references therein):

$$\lambda = \langle V_B - V_A \rangle_A - \Delta G_{AB} = \langle V_A - V_B \rangle_B + \Delta G_{AB} \qquad (a)$$

where ΔG_{AB} is the free energy difference between B and A, λ the corresponding reorganisation energy and $\langle \rangle_i$ denotes an average evaluated near the minimum of the potential i. Thus,

$$\Delta G_{AB} \simeq \tfrac{1}{2}(\langle \Delta V \rangle_A + \langle \Delta V \rangle_B) \qquad (b)$$

where ΔV now denotes the energy difference $V_B - V_A$. If the hydration of a single ion is considered, this can be shown to give $\Delta G^{el}_{sol} = \tfrac{1}{2}\langle V^{el}_{i\text{-}s} \rangle$, i.e. that the electrostatic contribution to the solvation energy equals half of the corresponding ion-solvent interaction energy (Warshel and Russell, 1984; Roux et al., 1990). Returning now to the inhibitor binding problem, this result may be exploited as indicated in Fig. 1. For each solvation process, i.e. solvation of the inhibitor in water and inside the protein, two states are considered where the first has the inhibitor molecule in vacuum and a non-polar cavity (given, e.g., by Lennard-Jones potential) already made in the given environment. The second state corresponds to the intact inhibitor molecule surrounded by

water or the solvated protein. The linear response approximation will then again give that $\Delta G^{el}_{bind} \simeq \frac{1}{2}\langle V^{el}_{i-s}\rangle$, where V^{el}_{i-s} is the solute-solvent electrostatic term. Hence, the electrostatic contribution to the binding free energy can be approximated by $\Delta G^{el}_{bind} \simeq \frac{1}{2}\langle V^{el}_{i-s}\rangle$ (where the Δ now refers to the difference between protein and water) and thus obtained from two MD simulations of the solvated inhibitor and of the inhibitor-protein complex.

The validity of the linear response results in the case of ionic solvation has been confirmed, e.g., in the study by Roux et al. (1990). Some additional calculations were also performed on simple systems that corroborate the approximation of equation b. These tests were carried out by comparing the free energy obtained from FEP/MD simulations of charging Na^+ and Ca^{2+} ions in a spherical water system (Åqvist, 1990) with the corresponding $\langle V^{el}_{i-s}\rangle$ from 75 ps MD trajectories. This yielded factors relating $\langle V^{el}_{i-s}\rangle$ to ΔG^{el}_{sol} of 0.49 for Na^+ and 0.52 for Ca^{2+}, both values being close to the predicted result of ½. A similar test on the charging of a methanol molecule, given by the OPLS potential (Jorgensen, 1986) in water gave a $\Delta G^{el}_{sol}/\langle V^{el}_{i-s}\rangle$ ratio of 0.43.

A crucial question was how to account for the contribution of non-polar interactions and hydrophobic effects to the free energy of binding which was termed ΔG^{vdw}_{bind}. In the ideal case, it should be possible to estimate this contribution from the non-polar (or van der Waals) interaction energies. The liquid theories of Chandler and coworkers (Chandler et al., 1983; Pratt and Chandler, 1977) have been successfully used to analyze hydrophobic effects and to calculate free energies of transfer for some non-polar molecules (Pratt and Chandler, 1977), but no analytical treatment of that kind seemed possible for solvation in an inhomogeneous environment such as a protein's active site. However, it was noted that the experimental free energy of solvation for various hydrocarbon compounds, such as n-alkanes, depends approximately linearly on the length of the carbon chain both in their own liquids

as well as in water (Ben-Naim and Marcus, 1984). MD simulations of n-alkanes solvated in water (Fig. 2) and in a non-polar van der Waals solvent (not shown) have been carried out, which indicate that also the average solute-solvent interaction energies vary approximately linearly with the number of carbons in the chain (the relationships being different in different solvents, of course). It would thus seem possible that a simple linear approximation of ΔG_{bind}^{vdw} from $\langle \Delta V_{i-s}^{vdw} \rangle$ might be able to account for the non-polar binding contribution. For instance, if σ is considered some appropriate measure of the size of the solute and if the solute-solvent van der Waals interaction energies and the corresponding non-polar free energy contributions (both in water and protein) depend linearly on σ, such that $\langle V_p^{vdw} \rangle = \alpha_p \sigma$, $\langle V_w^{vdw} \rangle = \alpha_w \sigma$, $\Delta G_p^{vdw} = \beta_p \sigma$ and $\Delta G_w^{vdw} = \beta_w \sigma$ then $\Delta G_{bind}^{vdw} = \frac{\beta_p - \beta_w}{\alpha_p - \alpha_w} \langle \Delta V_{i-s}^{vdw} \rangle$ is obtained. Since it seems difficult to derive a factor relating the two quantities in a reliable way from purely theoretical considerations, the approach was taken by the inventors to empirically try to determine such a relationship which is capable of reproducing experimental binding data. Thus, the free energy of binding is in one embodiment of the invention approximated by

$$\Delta G_{bind} = \tfrac{1}{2} \langle \Delta V_{i-s}^{el} \rangle + \alpha \langle \Delta V_{i-s}^{vdw} \rangle \qquad (1)$$

the parameter α being determined by empirical calibration.

Although, as discussed above, a theoretical prediction of the coefficient for $\langle \Delta V_{i-s}^{el} \rangle$ is ½, it may be practically useful to also treat this coefficient as an empirical parameter. This would lead to the free energy of binding being approximated by

$$\Delta G_{bind} = \beta \langle \Delta V_{i-s}^{el} \rangle + \alpha \langle \Delta V_{i-s}^{vdw} \rangle \qquad (1b)$$

where both parameters, α and β, are determined by empirical calibration.

DETAILED DISCLOSURE OF INVENTION

The assessment of the averages is normally performed by establishing 3-dimensional models or structural representations, using, e.g., suitable standard computer hardware and software, comprising a 3-dimensional structure of the receptor molecule alone and a 3-dimensional structure of the chemical substance "docked therein", and applying molecular dynamics calculations to the 3-dimensional representation. Force field data for use in the molecular dynamics calculations may be from any suitable force field such as publicly available force fields, e.g., AMBER (Weiner et al., 1986), CHARMM (Brooks et al., 1983), GROMOS, OPLS (Jorgensen, 1986), MM2 (Allinger, 1977), etc. The purpose of the molecular dynamics calculation is to be able to explore the available conformations of the system, thereby calculating the average interaction energies. In the simulation, the molecules are allowed to move around to enable exploration of the conformational space.

The establishment of the 3-dimensional structural representation may be performed using any method which will result in the establishment of the 3-dimensional coordinates of the molecule or combination in question, including crystallisation and X-ray crystallography, NMR, computer modelling, etc.

In the present context, the term "docked therein" indicates that the chemical substance has been brought to "fit" with the receptor molecule in the 3-dimensional representation; while this does not, in the present context, imply any numerical limitation with respect to the quality of the 3-dimensional fit, it is evident that the binding free energy resulting from the methods will reflect the degree of "fit". However, molecular dynamics applied to the system comprising the receptor molecule and the compound docked therein will in itself be useful for checking the correctness of the docking in case more than one position and/or more than one orientation is possible.

The calculation of the above-mentioned energy averages can, in principle, be performed separately, either manually or (preferably) by means of a computer. In both cases, the calculation may be performed on the basis of the well-known classical mechanical principles involving simulations on the equations of motion of the relevant molecular systems (molecular dynamics or MD); another principle could be the so-called Monte Carlo simulations, which does not actually solve the equations of motion, but rather calculate the probabilities of different conformations; a still further principle could be energy minimization in which the averages are replaced by minimum energies. Neither of the two last-mentioned principles are, however, contemplated to be as efficient as molecular dynamics. On the other hand, it should be understood that whenever reference is made to molecular dynamics in the following, the two other principles mentioned could also be interesting. It is preferred to perform the calculations using suitable software implementing the principle in question, preferably software which interfaces or communicates with the stored coordinates of the 3-dimensional models of the receptor molecule and of the receptor molecule with the chemical substance docked therein; such software is available as standard commercial software, e.g. one of the many commercially available types of computer software packages suitable for the purpose.

It will be understood that complex calculations involved in molecular dynamics simulations render the manual determination of the above-mentioned energy averages practically impossible, thus making the use of suitable computer soft- and hardware the preferred - and in practice necessary - choice.

It is one of the main features of the present invention that it has been found possible to arrive at a very high degree of conformity with actual measured values on the basis of such a simple procedure. The parameter with which the non-polar and hydrophobic energy difference, ΔV_{i-s}^{vdw}, is adjusted is suitably

a coefficient representing the result of a calibration established by comparing the results of predictions with actual measured values. The term "representing the result of a calibration" means that the parameter has been established by such calibration, or that the parameter has a value which would have resulted from such calibration against actual measured values; thus, it is not precluded that the parameter, although representing the results, could have been provided, in a particular case, in any other suitable manner.

It is believed that the adjusted parameter may be valid for a relatively broad range of systems, but when working in any particular system, it may be preferred to make a calibration against actual values measured on representatives of the system. Examples of such calibrations are given below. Thus, the calibration described in the below Example 1 was found to result in a value of the coefficient α of about 0.16. It is presumed that the numerical value of α will be at the most 1.0, and that most values of α in practice will be at the most 0.5 or preferably at the most 0.3, such as at the most 0.2. While these values are understood to be absolute values, it is believed that α will in fact be a positive value in most cases.

The coefficient for the electronic term $\langle \Delta V^{el}_{i\text{-}s} \rangle$ is predicted to be ½, but may also be treated as an empirical parameter (β) determined by calibration against known data. It is then believed that the parameter β will assume a value of at most 1.0, and that most values of β in practice will be about 0.2 - 0.8, such as about 0.3 - 0.7 or preferably about 0.4 - 0.6. As stated above, β will in a most preferred embodiment of the invention have a value of about 0.5. While these values are understood to be absolute values, it is believed (as is the case for the coefficient α) that β will in fact be a positive value in most cases.

In some systems, it seems suitable to add an additional constant term to Equation 1, so that the equation becomes

$$\Delta G_{bind} = \tfrac{1}{2}\langle \Delta V_{i-s}^{el} \rangle + \alpha \langle \Delta V_{i-s}^{vdw} \rangle + c \qquad (2)$$

where α is a coefficient representing the result of a calibration established by comparing the results of predictions with actual measured values (as described above), and c is a constant reflecting extrapolation to zero size of the chemical substance, that is, where the regression line is distinctly offset from origin when moving towards zero size of the chemical substance. The parameter c may also be used to correct for possible systematic errors due to e.g. the neglect of induced polarisation, possible force field deficiencies etc. In these cases, c will normally assume a value between -10 and 10 kcal/mol, typically between -3 and 3 kcal/mol, such as between -2 and 2 kcal/mol, e.g. between -1 and 1 kcal/mol. However, it is anticipated that in many cases, c can suitably be set to zero, as the extent of deviation will be of minor importance for the usefulness of the predicted values.

If also the electrostatic coefficient i treated as an empirical parameter, the present approximation assumes its most general form, namely

$$\Delta G_{bind} = \beta \langle \Delta V_{i-s}^{el} \rangle + \alpha \langle \Delta V_{i-s}^{vdw} \rangle + c \qquad (2b)$$

-where now both α, β and c are to be determined by empirical calibration.

While the solvent used in the above method is suitably and most often an aqueous solvent like water, it is within the scope of the invention to take any other suitable solvent as a starting point, including, e.g., methanol, ethanol, acetone, acetonitrile, chloroform, hexane, etc., or mixtures thereof or combinations of such solvents or mixtures thereof with water. The selection of the solvent will be of little importance to the predicted values as long as the solvent is one which is able to dissolve or solvate the receptor molecule and the substance (in the present context this means that

a sufficient amount of the receptor molecule can be homogeneously mixed with the solvent without precipitation so as to allow the determination of binding energies by some suitable method), but there may be cases where it is advantageous to modify the solvent environment (e.g. by modulating the ionic strength) in which the interaction of the substance and the receptor molecule is to take place. If the environment in which the interaction between the chemical substance, such as a drug, and a host or receptor molecule is to take place in the actual use of the drug is the human body, it might be particularly suitable to imitate e.g. human plasma as the solvent.

While the above discussion has primarily dealt with cases where the absolute free energy of binding is determined, the method of the invention also makes it possible to determine relative values of free energy of binding between a number of chemical substances capable of interacting with a host or receptor molecule.

As an example, the case may be considered where there are four inhibitors, I_1, I_2, I_3 and I_4. For each inhibitor, a $\langle \Delta V_{i-s}^{el} \rangle$ and a $\langle \Delta V_{i-s}^{vdw} \rangle$ are calculated from molecular dynamics simulations, or Monte Carlo simulations, or simply by energy minimization. For instance, for any particular guess of the parameter α, it can then be seen how good this guess is by comparing the calculated (from Equation 1) and observed values of ΔG_{bind} for each inhibitor. In order to find the best α, a least-squares optimization can be used, which means that the sum

$$\sum_{i=1}^{4} \left(\Delta G_{bind}^{calc} - \Delta G_{bind}^{obsd} \right)^2$$

is calculated and minimized. The minimization can be obtained by varying α systematically so that this sum reaches its minimum. The best α obtained in this way is thus the α that gives the least discrepancy between calculated and observed values (in what is called the least-squares sense).

Another approach is to determine α analytically. Using formula 1 and given that the absolute binding energies for n different inhibitors are known, α can be determined by minimizing the following sum of squares (SS) expression:

$$SS = \sum_{i=1}^{n} \left(\Delta G_{bind}^{calc} - \Delta G_{bind}^{obsd}\right)^2 = \sum_{i=1}^{n} \left(\frac{1}{2}<\Delta V_{i-s}^{el}> + \alpha<\Delta V_{i-s}^{vdw}> - \Delta G_{bind}^{obsd}\right)^2$$

As stated above, this can be done by iterative methods, but it is also possible to arrive at an exact result by partial differentiation with respect to α of the sum SS:

$$\frac{\partial SS}{\partial \alpha} = \sum_{i=1}^{n} \left[2<\Delta V_{i-s}^{vdw}> \cdot \left(\frac{1}{2}<\Delta V_{i-s}^{el}> + \alpha<\Delta V_{i-s}^{vdw}> - \Delta G_{bind}^{obsd}\right)\right]$$

and equalling the partially differentiated sum to zero, which yields:

$$\sum_{i=1}^{n} \left[2<\Delta V_{i-s}^{vdw}> \cdot \left(\frac{1}{2}<\Delta V_{i-s}^{el}> + \alpha<\Delta V_{i-s}^{vdw}> - \Delta G_{bind}^{obsd}\right)\right] = 0$$

$$\Updownarrow$$

$$\alpha = \frac{\sum_{i=1}^{n}\left(<\Delta V_{i-s}^{vdw}>\Delta G_{bind}^{obsd}\right) - \frac{1}{2}\sum_{i=1}^{n}\left(<\Delta V_{i-s}^{vdw}><\Delta V_{i-s}^{el}>\right)}{\sum_{i=1}^{n}<\Delta V_{i-s}^{vdw}>^2}$$

Accordingly, the parameters α and β and/or c can be determined by comparing the ΔG_{bind}^{calc} values (from any of the equations 1b, 2 and 2b) with ΔG_{bind}^{obsd} values and varying the parameters in the pertinent formula until a minimization of the sum of squares is obtained or by determining the parameters analytically by partial differentiation with respect to the parameters of the least squares expression. In the case formula 2b is used and an analytical solution is sought, the partial differentiation of the least square expression with respect to α, β and c will e.g. result in three linear equations with the three parameters to be determined.

The above-sketched procedures assume that there is access to the absolute observed values = ΔG_{bind}^{obsd}. Now, if for some

reason only the differences (relative values) of ΔG_{bind} are available for the four inhibitors, the actual ΔG_{bind} for each inhibitor would not be known, but the experimental data would consist of

$$\Delta\Delta G_{bind}\ (I_1\text{-}I_2) = \Delta G_{bind}\ (I_1) - \Delta G_{bind}\ (I_2)$$
$$\Delta\Delta G_{bind}\ (I_1\text{-}I_3) = \Delta G_{bind}\ (I_1) - \Delta G_{bind}\ (I_3)$$
$$\Delta\Delta G_{bind}\ (I_1\text{-}I_4) = \Delta G_{bind}\ (I_1) - \Delta G_{bind}\ (I_4)$$
$$\Delta\Delta G_{bind}\ (I_2\text{-}I_3) = \Delta G_{bind}\ (I_2) - \Delta G_{bind}\ (I_3)$$
$$\Delta\Delta G_{bind}\ (I_2\text{-}I_4) = \Delta G_{bind}\ (I_2) - \Delta G_{bind}\ (I_4)$$
$$\Delta\Delta G_{bind}\ (I_3\text{-}I_4) = \Delta G_{bind}\ (I_3) - \Delta G_{bind}\ (I_4)$$

There would thus be six experimental data points, each representing a difference in binding energy.

Again, for any guess of α (and β), the ΔG_{bind} values would be calculated (by one of the formulas 1 or 2) and their differences taken to obtain the $\Delta\Delta G_{bind}$ values. Then, in the same manner as above, the sum

$$\sum_{i=1}^{6} \left(\Delta\Delta G_{bind}^{calc} - \Delta\Delta G_{bind}^{obsd}\right)^2$$

would be minimized to obtain the optimum value of α (as well as β). Also in this case, the parameters α and β could be determined by iterative methods as well as analytically.

These parameters are then the ones that gives the best agreement with respect to the *relative* binding energies (or binding energy differences).

Accordingly, for a situation with n different inhibitors the sum to be minimized would then be

$$\sum_{i=1}^{N} \left(\Delta\Delta G_{bind}^{calc} - \Delta\Delta G_{bind}^{calc}\right)^2$$

-where $N = \frac{1}{2}(n^2-n)$, because there are $\frac{1}{2}(n^2-n)$ differences in binding energy between inhibitors in a system with n inhibitors.

One could also imagine a case where, for some reason, the calculations give a systematic error in the absolute ΔG_{bind} values. It might then be desirable to try to obtain the best fit of the relative energies instead, and the latter type of the above-described optimization methods would then be preferable.

It will be evident to the person skilled in the art that the method of the invention can be utilized in a number of ways to shorten drug development time and make molecular modelling more efficient. The most evident advantage seems to be that considerable synthesis efforts and thereby very considerable time- and resource-consuming activities can be saved on the way towards realistic drug candidates. On the other hand, as will be explained in greater detail herein, because the method of the invention involves an empirical element, the possibilities of performing relevant calibrations and experimental confirmation should not be neglected where such calibration or confirmation is indicated.

A couple of scenarios can be given to illustrate the possibilities provided by the invention:

1) When considering a new system (a new macromolecule for which it is desired to design a drug interacting with the macromolecule), the initial stage could be to establish a possible lead compound by some means, e.g. computer modelling. The method of the invention could then be applied, for example using equation 1 with a value of α established earlier (for some other system, e.g. the value of α of about 0.16 disclosed herein). If the outcome of the calculations according to the method of the invention would indicate that the binding of the lead compound is good enough, the lead compound could be synthesized, and its actual binding power could be measured. If the outcome of the calculations do not indicate that the binding of the lead compound is good enough, a next stage could be to try to modify the lead

compound "manually", using computer graphics, in order to improve the binding, and then perform the method of the invention on the modified compound; these steps could be repeated one or several times until a compound had been designed the data of which would show that the compound would be worth synthesizing. As mentioned above, it is not unlikely that even for a completely new system, a suitable drug candidate can be found without any recalibration of α.

2) If, after comparing the results of the method of the invention for a series of compounds with actual measured values for a new system, it is found that the agreement using an earlier established value of α is not satisfactory, a recalibration would be performed on the system in question, and then the further procedure would be as in 1) above.

3) If it is found that the calibration of α is not "universal", a number of α's characterizing different systems could be established. It is contemplated that in such a case, a given α would pertain to a certain class of (chemically similar) inhibitors or (less likely) a class of e.g. similar protein binding sites, or a combination of these two possibilities. A change of the solvent from e.g. water to methanol may change the optimal value of α. In any case, for a new system, one would then choose an earlier established value of α for a similar system.

Of course, when using any of the formulas 1b, 2, and 2b it is necessary to provide suitable values for both α and β and/or c, for example by the methods outlined above.

While several other scenarios can be envisaged, it is believed that one important utilization of the method of the invention will be a method for identifying a chemical substance capable of interacting with a host or receptor molecule, e.g. binding to the host or receptor molecule, with a

predicted binding energy equal to or better than a predetermined threshold value, comprising

1) choosing a chemical substance, A, which could potentially interact with the host or receptor molecule, and providing a 3-dimensional structural representation thereof,

2) predicting the binding free energy between the chemical substance A and the host or receptor molecule by the method of the invention,

3) if the predicted binding free energy between the chemical substance A and the host or receptor molecule determined is equal to or better than the predetermined threshold value, then identifying the chemical substance A as the chemical substance X,

4) if the predicted binding free energy between the chemical substance A and the host or receptor molecule determined is not equal to or better than a predetermined threshold value, modifying the 3-dimensional structural representation and predicting the binding free energy between the thus modified chemical substance, B, and the host or receptor molecule by the method of the invention, and

5) if necessary repeating step 4 until the predicted binding free energy determined between the resulting chemical substance, X, and the host or receptor molecule is equal to or better than the predetermined threshold value.

One suitable way of calibrating the above-mentioned method could comprise providing a sample of a chemical substance or chemical substances selected from the chemical substances A, B and X and providing a sample of the host or receptor molecule, measuring the binding free energy between the chemical substance or substances and the host or receptor molecule, and if the measured binding free energy between the chemical substance or substances and the host or receptor

molecule is not equal or substantially equal to the predicted value, then performing a calibration of the method according to the invention so as increase the predictive value of the method.

LEGEND TO FIGURES

Fig. 1. Illustration of how the approximation of equation b can be used to estimate the electrostatic contribution to ΔG_{sol} in a given environment that can either be water or protein. The state A corresponds to a system with the isolated inhibitor in the gas-phase and a ready-made van der Waals cavity in the condensed system. State B is simply the solvated inhibitor in water or in the protein's active site. These states are given by the two potentials V_A and V_B. For any given configuration of the system, equation b will reduce to $\frac{1}{2}V_{i-s}^{el}$ since the solvent configuration in state A is uncorrelated with the charge distribution of the solute in state B.

Fig. 2. Calculated dependence of the average solute-solvent interaction energy on the length of the carbon chain for n-alkanes solvated in water.

Fig. 3. Chemical structures of the inhibitors I_1, ..., I_5 used in the calculations in Example 1.

Fig. 4. Stereo view of the 50 ps average MD structure of the EP-I_1 complex (thin lines) superimposed on the corresponding crystallographic structure (see Example 1). The active site of the enzyme is shown with the bound inhibitor in the centre of the picture.

EXAMPLE 1

As a test system, endothiapepsin (EP) was chosen; it belongs to the family of aspartic proteinases (see, e.g., Davies, 1990; Fruton, 1976), a class of enzymes for which numerous studies of inhibitor binding have been reported. Crystal

structures of native endothiapepsin and inhibitor complexes have been published by Blundell and coworkers (Foundling et al., 1987; Veerapandian et al., 1990). In the present Example, the crystal structure of endothiapepsin was used in complex with the inhibitor H218/54 (I_1, Fig. 3), recently determined by Symbicom AB, as the structural starting point for the computations. Four other inhibitor compounds were used in this work and their chemical structures are also shown in Fig. 3 (I_2-I_5). Experimental binding data for the inhibitors have been obtained (see A.K. Falknäs, Graduation report, University of Gothenburg, Sweden).

Computational details

Starting from the experimental structure of the EP-I_1 complex, the other inhibitors were "manually" built into the EP active site using the FRODO (Jones, 1978) and Hydra (Hubbard, 1984) graphics software. This model building was relatively straightforward since all the compounds contain the invariant hydroxyethylene transition state isostere adjacent to a peptide group, also present in the inhibitors studied by Blundell's group (cf. above). After initial energy minimisation, MD simulations consisting of 25 ps of equilibration and 50 ps of data collection were performed for each of the inhibitors, solvated in water as well as bound to the solvated protein.

The ENZYMIX programme (Warshel and Creighton, 1989) was used for all MD simulations together with the GROMOS potential (van Gunsteren and Berendsen, 1987). This force field was revised by the present inventors with respect to the oxygen-extended carbon (CH_n) interactions, since the standard GROMOS parameters were found from FEP simulations to give an incorrect value of ΔG_{hydr} = +0.1±0.5 kcal/mol for methane. By changing the oxygen repulsive Lennard-Jones 6/12 parameter (A_o) for O-CH_n interactions from 421.0 to 793.3 (kcal/mol)^{½}Å^6, ΔG_{hydr} for methane becomes 2.5±0.4 kcal/mol, in better agreement with the experimental value of 2.0 kcal/mol

(Ben-Naim and Marcus, 1984). This revision is quite important for a correct description of the hydrophobic effect. Also the Lennard-Jones parameters for charged carboxylate groups had been revised in order to better reproduce experimental solvation energies (Åqvist et al., 1993). The present calculations used a spherical water "droplet" of radius 16 Å for the simulations of inhibitors in solution and a corresponding 16 Å sphere containing both protein atoms and water in the simulations of bound inhibitors. The atoms within this sphere were free to move while protein atoms beyond 16 Å were restrained to their crystallographic positions. An interaction cutoff radius of 8 Å was used and the MD time step was 0.001 ps. The equilibration phase of the protein simulations consisted of 5 ps of successive heating of the system and weakening of harmonic positional restraints that were applied to the protein atoms. After this period all restraints within the 16 Å sphere were set to zero and the system was further equilibrated at the final temperature of 298 K for another 20 ps. The r.m.s. coordinate deviation for the inhibitor atoms between the following 50 ps MD average structure and the experimental EP-I_1 complex is 0.94 Å. Fig. 4 shows a superposition of these two structures, and it can be seen that the agreement is quite satisfactory.

Results and discussion

The first four inhibitors (I_1, ..., I_4) were used as the calibration set in order to determine the optimum value of the parameter α and examine the success of eq. 1 in reproducing the experimental results. Table 1 shows the observed and calculated *absolute* free energies of binding for the inhibitors 1_1, ..., I_4 using the value of $\alpha = 0.161$ which optimizes the r.m.s. agreement with experiment.

Table 1

Calculated and observed absolute binding free energies for inhibitors I_1, ..., I_4 to endothiapepsin (in kcal/mol).

I_x	ΔG_{calc}	ΔG_{obsd}
1	-11.16	-10.69
2	-8.23	-7.93
3	-11.14	-11.67
4	-6.29	-6.57

The above-mentioned value, 0.161, of α gives a mean unsigned error of 0.39 kcal/mol for the calibration set and the largest error is 0.53 kcal/mol for inhibitor I_3. Such an accurate fit had not been expected for the absolute binding energies; the initial main purpose was to be able to apply and calibrate equation 1 with respect to the relative ones. If the calibration is instead done using relative binding energies, one obtains a very similar value of α (0.169) with mean unsigned errors that are virtually identical to those above. This result is quite remarkable and thus lends further support to the method of the invention, embodied in the approximation of equation 1. The calculated and observed *relative* free energies of binding (using α = 0.161) are shown in Table 2 and the average unsigned error is 0.59 kcal/mol.

Table 2

Calculated and observed relative binding free energies for inhibitors I_1, ..., I_4 to endothiapepsin (in kcal/mol).

I_x-I_y	$\Delta\Delta G_{calc}$	$\Delta\Delta G_{obsd}$
1-2	-2.92	-2.76
1-3	-0.01	+0.98
1-4	-4.87	-4.12
2-3	+2.91	+3.74
2-4	-1.95	-1.36
3-4	-4.86	-5.10

The largest error here is 0.99 kcal/mol for the I_1/I_3 selectivity, but all other relative free energies have the correct sign and are within 0.8 kcal/mol of the experimental result. It is particularly interesting to note that the calculations were able to discriminate the low-affinity inhibitor I_4 quite well from the high-affinity ones (I_1 and I_3).

As the above results were encouraging, the predictive power of the approach was tested by modelling an inhibitor not present in the calibration set. For this, the inhibitor I_5 was chosen; as can be seen from Fig. 3, this inhibitor differs significantly in its chemical structure from any member of the calibration set. This molecule was built into the EP active site and subjected to the same simulation procedure as the other inhibitors. The predicted absolute free energy of binding for I_5 is -9.70 kcal/mol which is in excellent agreement with the corresponding observed result of $\Delta G_{bind}(I_5)$ = -9.84 kcal/mol. The calculated relative binding free energies with respect to the four inhibitors in the calibration set are given in Table 3, where it can be seen that all the pairwise selectivities involving I_5 are correctly predicted by the simulations, the maximum error in this case being 0.61 kcal/mol.

Table 3

Calculated and observed relative binding fee energies involving inhibitor I_5 (in kcal/mol).

I_x-I_y	$\Delta\Delta G_{calc}$	$\Delta\Delta G_{obsd}$
1-5	-1.46	-0.85
2-5	+1.47	+1.91
3-5	-1.44	-1.83
4-5	+3.41	+3.27

The above results indicate that the simple linear approximation of the free energy according to the method of the invention is able to describe the main physics (electrostatic

(polar) and hydrophobic (non-polar) interactions) of the binding process in a satisfactory way.

It is, of course, important to try to identify and separate the different types of errors involved in the present method and suggest how they could possibly be dealt with. There are basically four sources of errors, namely (i) inaccurate starting structures due to incorrect model building, (ii) possible deficiencies in potential energy functions, (iii) poor MD convergence, e.g., due to short trajectories, and (iv) the approximation of eq. 1 itself. The only remedy for errors of the first type is to obtain as much structural information as possible, and in order to assess their magnitude it would be desirable to carry out calculations on a set of inhibitor complexes whose 3-D structures all have been experimentally determined. Although MM/MD potentials are continuously being refined by many research groups, errors of the second type may still be considerable. Thus, for example, it is not yet entirely clear how well customary protein force fields actually reproduce relevant energetic properties. In the present case, it can be noted that the force field used above required some revision in order to reproduce essential solvation properties (see above).

The convergence properties of MD stimulations depend on how far from equilibrium the initial structure is, but judging from the present Example, it seems that one can reach satisfactory convergence within reasonable computing time. For example, by comparing averages over the first and second halves of the MD trajectories, average (over all five inhibitors) errors were obtained of ±0.35 and ±0.75 kcal/mol for V_{i-s}^{vdw} and V_{i-s}^{el}, respectively, in the protein; the corresponding errors in water are ±0.46 and ±0.62 kcal/mol. This would yield a nominal error range of ±0.82 kcal/mol in equation 1 originating from the MD convergence uncertainty. Here, it can be noted that it seems to be an important advantage of the method of the present invention that it focuses on simulation of the thermodynamically relevant states compared,

e.g., to Free Energy Perturbation calculations where most of the computing time is spent on the paths between such states. The errors associated with the approximation of eq. 1 are difficult to estimate quantitatively although the agreement with experimental binding data obtained here indicates that the linear approximation is reasonable. A larger calibration set could obviously be desirable, but this is mainly a matter of refinement of this particular embodiment of the main principle of the invention. As mentioned above, Ben-Naim and Marcus (1984) have shown for several classes of hydrocarbon chain containing organic compounds that a linear fit of $\Delta G_{sol} = kn + l$, where n is the number of carbon atoms in the chain, is quite accurate both in water and non-polar solvents. It can, however, be noted that in some cases the extrapolation, l, of ΔG_{sol} to zero chain length is non-zero (e.g., for n-alkanes in water $l \simeq 1.42$ kcal/mol). As discussed further above, this might suggest that an additional constant term should be added to eq. 1, reflecting a difference in extrapolation to "zero inhibitor size" (between the water and protein environments) of the non-polar contribution to ΔG_{sol}. It is also important to emphasize that equation 1 with the above parameterisation of α is *not* an equation for the individual solvation energy terms, since the factor α represents the combined effect of several energy/free energy relations. The fact that solvation energy differences are always dealt with may also cause some cancellation of possible systematic errors. Furthermore, due to the larger weight given by equation 1 to the V^{el}_{i-s} term, the formula might be expected to give better results for polar inhibitors, as long as the electrostatic linear response approximation holds. However, for inhibitors carrying a net charge, long range electrostatic effects as well as induced polarisation effects are known to be important, but these problems have not yet been resolved in most available MD programs and force fields. When comparing inhibitors of different charge states long-range corrections of the Born type (see, e.g., Straatsma and Berendsen, 1988; Åqvist, 1990) would obviously

become important and may make accurate predictions more difficult.

As mentioned further above, it is possible that the empirical parameter α is readily transferable to other systems, but it is also possible that it will display some system dependency. Given the fact that the parameterisation of force fields can differ considerably, α may be found to be force field specific.

However, the above detailed considerations concerning possible refinements of the specific embodiments disclosed herein do not detract from the very important fact that the new overall principle according to the present invention is a very considerable advance in the art of computational chemistry and molecular modelling, making it possible to perform drug design tasks with considerably less resource consumption and in considerably shorter time than hitherto.

EXAMPLE 2

The aspartic proteinase from the human immunodeficiency virus type 1 (HIV-1) is the target of intense AIDS drug development. This enzyme and three of its inhibitors were chosen as a second test system (Hansson, 1994).

Computational details

Starting coordinates were from the 5HVP entry in the Brookhaven Protein Data Bank (Fitzgerald et al. 1990), describing a complex of the proteinase with the inhibitor acetylpepstatin, designated Inhibitor 1 in this example.

Most charges of the protein were turned off, by replacing charged groups by dipoles with zero total charge. Charges left on were as follows (numbering of 5HVP).

AA↓	State→	Neut	$Cent^-$	$CentH_2$	$Cent^{2-}$
Asp	25	-	-	H	-
Asp	225	H	H	H	-
Arg	8	+	+	+	+
Arg	208	+	+	+	+
Asp	29	-	-	-	-
Asp	229	-	-	-	-
Asp	30	-	-	-	-
Asp	230	-	-	-	-
Glu	34	-½	0	0	0
Glu	234	-½	0	0	0
Lys	45	+	0	0	0
Lys	245	+	0	0	0
Arg	87	+	+	+	+
Arg	287	+	+	+	+

The 27 water molecules closest to the active site of the protein were retained from the X-ray coordinates file. Actually only one water molecule (water 308H in 5HVP) is very near the centre, and the rest form the proximal parts of the water shell around the protein. This central water (308H) is in a conserved position for a water molecule in all early structures reported for HIV-1 proteinase (Fitzgerald and Springer 1991). It was retained in the calculations for Inhibitor 1 and Inhibitor 2, but was removed in the calculations for Inhibitor 3 due to the fact that Inhibitor 3 is engineered to mimic this water molecule with its protruding carbonyl oxygen.

Inhibitor 1: acetyl-pepstatin

This is the N-capped oligopeptide Ace-Val-Val-Sta-Ala-Sta, where Sta is the rare amino acid residue *statine* (Sta, (4*S*,3*S*) 4-amino 3-hydroxy-6-methyl-heptanoic acid), and Ace

is an acetyl group: $CH_3(CO)$ (Richards et al., 1989). Coordinates for Inhibitor 1 were taken from 5HVP. The C-terminal carboxyl group is negatively charged.

Inhibitor 2: pepstatin

This is the N-capped oligopeptide Iva-Val-Val-Sta-Ala-Sta, where Iva is an isovaleryl group; $(CH_3)_2CHCH_2(CO)$. The only difference from Inhibitor 1 is the added group of three nonpolar carbon atoms (isopropyl group) in the beginning of the molecule (Dreyer et al. 1989). Inhibitor 2 was modeled from Inhibitor 1 by adding the 3 carbon atoms, and rotating around the bonds at the end of the molecule to remove any collisions between atoms.

Inhibitor 3: DMP 323

This is the substituted seven-membered cyclic urea DMP 323 (Lam et al., 1994), developed by the DuPont Merck Pharmaceutical Company. The coordinates for the complex of Inhibitor 3 and the proteinase had not yet been deposited into the Protein Data Bank. Therefore this inhibitor was model built: simulated annealing was used to determine a low energy structure of the inhibitor and this was modelled into the active site of the proteinase (5HVP coordinates).

Simulation parameters

A *sphere* of radius 20 Å of water was added, and any protein atoms outside this sphere were kept fixed. The *time step* was 1 fs, except for some equilibration steps where noted (2 fs). The *temperature* was 300 K.

The *cutoff distance* was 15 Å in the first set of calculations for all three inhibitors. Then, Inhibitor 3 was tested with 15 Å, 10 Å and 8 Å. Tests of the turning off of charges were performed using 10 Å, as were the pH tests described below. A 10 Å cutoff was used for the FEP calculations where Inhibitor

2 was changed into Inhibitor 1. These cutoffs apply to interactions between protein groups of zero net charge. Charged protein groups, and all parts of the inhibitor, interact with all other parts of the simulated system, without cutoffs.

The molecular dynamics simulations were performed using the program ENZYMIX and the GROMOS potential, with modifications as in Example 1.

Calculation of binding energies

The three inhibitors were simulated, with a 15 Å cutoff, until values were reasonably stable. The protein was electrically neutral ('Neut'). The approximation of equation (1) was applied, with α = 0.161 taken from Example 1.

Inhibitor 1:

$$\Delta G_{bind} \approx \tfrac{1}{2}\cdot((-255.84)-(-255.19))+0.161\cdot((-89.13)-(-47.29))$$
$$= (-0.32)+(-6.74) = \mathbf{-7.1\ kcal/mol}$$

Inhibitor 2:

$$\Delta G_{bind} \approx \tfrac{1}{2}\cdot((-233.01)-(-234.99))+0.161\cdot((-100.26)-(-55.08))$$
$$= (+0.98)+(-7.27) = \mathbf{-6.3\ kcal/mol}$$

Inhibitor 3:

$$\Delta G_{bind} \approx \tfrac{1}{2}\cdot((-75.02)-(-69.14))+0.161\cdot((-89.86)-(-43.39))$$
$$= (-2.94)+(-7.42) = \mathbf{-10.4\ kcal/mol}$$

Table 4

Inhibitor 1: Acetyl-pepstatin					
15 Å		time ps	V^{vdw} kcal/mol	V^{el} kcal/mol	V^{pot} kcal/mol
PROTEIN (Neut):	Equilibration	110.2			
	Data collection	25.0	-89.05	-256.02	-6073.69
		25.0	-91.47	-253.48	-6063.06
		25.0	-89.50	-251.68	-6037.51
		25.0	-86.00	-264.39	-6029.77
		7.6	-87.38	-267.00	-6012.86
		17.4	-90.63	-247.79	-6028.66
	Average	125.0	-89.13	-255.84	
Water:	Equilibration	130.0	(2fs)		
	Data collection	25.0	-46.26	-257.22	-10492.30
		25.0	-47.65	-251.52	-10512.25
		25.0	-47.53	-251.01	-10502.57
		25.0	-47.66	-261.85	-10517.91
		25.0	-47.36	-254.38	-10519.44
	Average	125.0	-47.29	-255.19	

Table 5

Inhibitor 2: Pepstatin					
15 Å		time ps	V^{vdw} kcal/mol	V^{el} kcal/mol	V^{pol} kcal/mol
PROTEIN (Neut):	Equilibration	85.2			
	Data collection	25.0	-103.41	-225.95	-6078.87
		25.0	-100.77	-231.47	-6161.13
		25.0	-98.90	-240.61	-6168.42
		25.0	-99.59	-237.60	-6149.30
		25.0	-98.61	-229.40	-6126.96
	Average	125.0	-100.26	-233.01	
Water:	Equilibration	138.0	(2fs)		
	Data collection	25.0	-54.51	-237.80	-10471.33
		25.0	-55.44	-234.92	-10472.92
		25.0	-55.99	-224.29	-10469.69
		25.0	-55.94	-235.67	-10470.89
		25.0	-53.54	-242.29	-10478.81
	Average	125.0	-55.08	-234.99	

Table 6

Inhibitor 3: DMP 323					
15 Å		time ps	V^{vdw} kcal/mol	V^{el} kcal/mol	V^{pot} kcal/mol
PROTEIN (Neut):	Equilibration	85.2			
	Data collection	25.0	-89.39	-76.42	-5914.07
		25.0	-89.60	-74.59	-5919.20
		9.8	-90.05	-72.76	-5919.46
		15.2	-88.96	-74.87	-5916.97
	Average	75.0	-89.46	-75.02	
Water:	Equilibration	50.0			
	Data collection	25.0	-43.51	-67.88	-10405.85
		25.0	-43.99	-70.02	-10417.31
		25.0	-42.67	-69.53	-10412.99
	Average	75.0	-43.39	-69.14	

Cutoff effect

For Inhibitor 3, three different cutoff distances were tried, 15 Å, 10 Å And 8 Å. Of these, the first two converged properly, and gave similar results. The 8 Å run did not converge, and the coordinates showed that some charged amino acids had taken on a very different conformation, exposing themselves much more to the surrounding water. This demonstrates the effect of overpolarization of water, where a short cutoff gives the water overly polar properties, making it too energetically favourable for charged groups to be exposed to water. The following result was obtained for the 10 Å run.

Table 7

Inhibitor 3: DMP 323; 10 Å cutoff					
10 Å		time ps	V^{vdw} kcal/mol	V^{el} kcal/mol	V^{pot} kcal/mol
PROTEIN (Neut):	Equilibration	60.2			
	Data collection	25.0	-85.76	-76.64	-5909.37
		25.0	-85.96	-78.20	-5896.49
		25.0	-85.73	-77.56	-5905.23
	Average	75.0	-85.82	-77.47	
Water:	Equilibration	50.0			
	Data collection	21.5	-42.88	-72.60	-10423.32
		3.5	-44.33	-60.54	-10420.80
		25.0	-43.78	-70.17	-10437.82
		25.0	-43.82	-65.82	-10435.30
	Average	75.0	-43.56	-68.97	

Inhibitor 3: 10 Å

$$\Delta G_{bind} \approx \tfrac{1}{2}\cdot((-77.47)-(-68.97))+0.161\cdot((-85.82)-(-43.56))$$
$$= (-4.25)+(-6.80) = \mathbf{-11.0\ kcal/mol}$$

The result was sufficiently similar to justify using 10 Å cutoffs in the following simulations.

Electrical effects

The amino acid residues far from the inhibitor had their charges turned off. On the other hand, charges close to the inhibitor were left on.

To investigate the effects of turning off charges of amino acid residues at 'middle distance' from the inhibitor, a different configuration was tested, where Glu 34, Glu 234, Lys 45 and Lys 245 were turned off. The same water simulation is used for both protein configurations. Therefore, relative

effects of charges to the protein can be studied by comparing protein simulations only.

Table 8

Inhibitor 3: DMP 323; 10 Å cutoff					
10 Å		time ps	V^{vdw} kcal/mol	V^{el} kcal/mol	V^{pot} kcal/mol
PROTEIN (Neut):	Equilibration	60.2			
	Data collection	25.0	-87.51	-80.52	-5767.96
		25.0	-88.97	-73.78	-5791.21
		24.0	-89.25	-77.71	-5783.69
	Average	74.0	-88.57	-77.33	
Water:	Equilibration	60.2			
	Data collection	25.0	-85.76	-76.64	-5902.37
		25.0	-85.96	-78.20	-5896.49
		25.0	-85.73	-77.56	-5905.23
	Average	75.0	-85.82	-77.47	

The electrical term almost did not change. The result justifies using the 'Cent$^-$' charges for other calculations on small uncharged inhibitors like Inhibitor 3. These four peripheral charges where turned off in the following pH studies.

Effects of pH

The binding of inhibitors to HIV-1 proteinase is pH dependent (Hyland et al., 1991, Richards et al., 1989). Generally, pH below 6 gives better binding than pH above 7. This may stem from pH-induced conformational changes, but the protonation state of the two active site aspartates should also be involved. These have a pK_a split from sitting close to one another and therefore have pK_a's in the ranges 3.1-3.7 and 5.2-6.5, respectively (Hyland et. al, 1991).

To test the effect of protonation of the active site Asp 25 and Asp 225 on binding of Inhibitor 3, the two protein states '$Cent^{2-}$' and '$CentH_2$' were compared to the '$Cent^-$' calculations. The same water can be used.

Table 9

Inhibitor 3: DMP 323; protonation of central Asp's					
10 Å		time ps	V^{vdw} kcal/mol	V^{el} kcal/mol	V^{pot} kcal/mol
PROTEIN (Neut):	Equilibration	60.2			
	Data collection	25.0	-87.51	-80.52	-5767.96
		25.0	-88.97	-73.78	-5791.21
		24.0	-89.25	-77.71	-5783.69
	Average	74.0	-88.57	-77.33	
Water:	Equilibration	60.2			
	Data collection	25.0	-86.40	-74.00	-5788.97
		25.0	-86.81	-75.48	-5789.14
		25.0	-87.44	-75.54	-5817.64
	Average	75.0	-86.88	-75.01	
PROTEIN ($CentH_2$):	Equilibration	60.2			
	Data collection	25.0	-90.00	-61.93	-5658.56
	Average	25.0	-90.00	-61.93	

These calculations predict low binding affinity at very low pH when both aspartates are protonated ('$CentH_2$'). The best binding is found in the singly protonated state ('$Cent^-$'), but the doubly negative (high pH) state ('$Cent^{2-}$') is predicted to give fair binding also.

Free Energy Perturbation

A classical free energy perturbation was performed, changing Inhibitor 2 into Inhibitor 1, by removing three carbon atoms.

- First the atoms were shrunk, while atom distances were constrained to normal bond lengths.

- Then the atoms were pulled in to 20% of their normal bond length, while at the same time shrinking them more. This means pulling the three now small atoms inside the end-carbon of Inhibitor 1.

- At last the atoms were allowed to disappear completely.

The protein was in the 'Neut' electric state. Protein simulations started from the end of the 210 ps run reported above for inhibitor 2, and the water simulations started from the end of the corresponding 263 ps run.

Two calculations were run, with 1 ps and 2 ps per λ-point. A total of 68 λ-points were used for the protein and 61 for the water calculation. The first step of the three took more points to converge in the protein (40) than in water (33). The second two steps were book-kept together, 28 points in either case. The first 25% of the data collected from each point was discarded. Errors in the table below (table 10) are hysteresis differences.

Table 10

1 ps/step	PROTEIN		WATER	
Inhibitor 2 ↓ Intermediate	40 × 0.75 ps	-0.94 ± 0.16	33 × 0.75 ps	0.09 ± 0.20
↓ Inhibitor 1	28 × 0.75 ps	-8.69 ± 0.08	28 × 0.75 ps	-9.21 ± 0.14
TOTAL:	68 × 0.75 ps	-9.63 ± 0.24	61 × 0.75 ps	-9.12 ± 0.34
DIFFERENCE:	-0.51 ± 0.58			
2 ps/step	PROTEIN		WATER	
Inhibitor 2 ↓ Intermediate	40 × 1.50 ps	0.27 ± 0.15	33 × 1.50 ps	-0.52 ± 0.16
↓ Inhibitor 1	28 × 1.50 ps	-8.55 ± 0.14	28 × 1.50 ps	-7.06 ± 0.32
TOTAL:	68 × 1.50 ps	-8.28 ± 0.29	61 × 1.50 ps	-7.58 ± 0.48
DIFFERENCE:	-0.70 ± 0.77			

The result of the calculation is that Inhibitor 1 binds 0.7 kcal/mol tighter than Inhibitor 2. There is a difference between 1 ps and 2 ps calculations, but not so large as to make an even longer calculation worthwhile.

Agreement between methods

The values calculated by equation (1) were consistent with the free energy perturbation. The method of the invention predicts that Inhibitor 1 will bind 0.8 kcal/mol better than Inhibitor 2. The FEP calculation gibes 0.7 kcal/mol - an excellent agreement.

Comparison with experimental data

Experimental values (Dreyer et al., 1989, Lam et al., 1994, Richards et al., 1989) on the binding of the three inhibitors are:

Table 11

Inhibitor	pH	T K	K_i nM	ΔG_{bind} kcal/mol	$\Delta G_{bind}^{adjusted}$ kcal/mol	ΔG_{calc} kcal/mol	ΔG_{diff} kcal/mol
1	5.0	303	35	-10.2	-10.2	-7.1	3.1
2	6.0	310	1400	-8.0	-8.7	-6.3	2.4
3	5.5	298	0.27	-13.1	-13.1	-10.4	2.7
						-11.0	2.1

$\Delta G_{bind}^{adjusted}$ is the experimental value, adjusted if the measurement was made at high pH by adding the free energy needed to reach the singly protonated form (a pKa of 5.5 for the upper aspartate is assumed). The rationale behind this adjustment is the reported (Richards et al., 1989) very poor binding of Inhibitor 1 at pH 7.0, suggesting that also Inhibitor 2 binds to the singly protonated form of the protein.

The approximation of equation (1) gives the same relative order of binding strength for the three inhibitors as experimentally observed:

Inhibitor 3 > Inhibitor 1 > Inhibitor 2

where '>' means 'binds better than'. The FEP calculation also gives the proper relation between inhibitors 1 and 2.

The approximate calculations consistently lack 2 to 3 kcal/mol. This error could of course be minimized by determination of the optimum value of c as described above.

REFERENCES

- Allinger N L (1977), *J. Amer. Chem. Soc.* **99**, 8127.
- Åqvist J (1990), *J. Phys. Chem.* **94**, 8021-8024.
- Åqvist J, Fothergill M and Warshel A. (1993), *J. Am. Chem. Soc.* **115**, 631-635.
- Bash P A *et al.* (1987), *Science* **235**, 574-576.

- Ben-Naim A and Marcus Y (1984), *J. Chem. Phys.* **81**, 2016-2027.
- Beveridge D L and DiCapua F M (1989), *Ann. Rev. Biophys. Biophys. Chem.* **18**, 431-492.
- Brooks B R *et al.* (1983) *J. Comput. Chem.* **4**, 187.
- Brooks C L III (1989), in: van Gunsteren W F and Weiner P K (eds.), *Computer Simulation of Biomolecular Systems*. ESCOM, Leiden, 73-88.
- Chandler D, Weeks J D and Andersen H C (1983), *Science* **220**, 787-794.
- Cohen N C *et al.* (1990), *J. Med. Chem.* **33**, 883-894.
- Davies D R (1990), *Ann. Rev. Biophys. Biophys. Chem.* **19**, 189-215.
- Dixon J S (1992), *TIBTECH* **10**, 357-363.
- Dreyer G B *et al* (1989), *Proc. Natl. Acad. Sci. USA*, **86**, 9752-9756.
- Fitzgerald P M D *et al.* (1990), *J. Biol. Chem.*, **265**, 14209-14219.
- Fitzgerald P M D and Springer J P (1991), *Annu. Rev. Biophys. Biophys. Chem.*, **20**, 299-320.
- Foundling S I *et al.* (1987), *Nature* **327**, 349-352.
- Fruton J S (1976), *Adv. Enzymol.* **44**, 1-36.
- Hansson T (1994), *'Filosofie magister (biologi)-degree project report, Uppsala University'*.
- Hubbard R E (1984), in: *The Representation of Protein Structure, Proc. Computer-Aided Molecular Design Conference*. Oyez, New York, p. 99.
- Hyland L J, Tomaszek Jr. T A and Meek T D (1991), *Biochemistry*, **30**, 8454-8463.
- Jones T A (1978), *J. Appl. Cryst.* **11**, 268-272.
- Jorgensen W L (1986), *J. Phys. Chem.* **90**, 1276-1284.
- Jorgensen W L (1989), *Acc. Chem. Res.* **22**, 184-189.
- Lam P Y S *et al.* (1994), *Science*, **263**, 380-384.
- Lee F S *et al.* (1992), *Prot. Eng.* **5**, 215-228.
- Marcus R A (1964), *Ann. Rev. Phys. Chem.* **15**, 155-196.
- McDonald J J and Brooks C L III (1991), *J. Am. Chem. Soc.* **113**, 2295-2301.

- Merz K M Jr. and Kollman P A (1989) *J. Am. Chem. Soc.* **111**, 5649-5658.
- Merz K M Jr., Murcko M A and Kollman P A (1991), *J. Am. Chem. Soc.* **113**, 4484-4490.
- Mitchell M J and McCammon J A (1991), *J. Comp. Chem.* **12**, 271-275.
- Pearlman D A and Kollman P A (1991), *J. Chem. Phys.* **94**, 4532-4545.
- Pratt L R and Chandler D (1977), *J. Chem. Phys.* **67**, 3683-3704.
- Rao B G and Singh U C (1991), *J. Am. Chem. Soc.* **113**, 6735-6750.
- Richards A D *et al.* (1989), *FEBS Lett.*, **247**, 113-117.
- Roux B, Yu H-A and Karplus M (1990), *J. Phys. Chem.* **94**, 4683-4688.
- Straatsma T P and Berendsen H J C (1988), *J. Chem. Phys.* **89**, 5876-5886.
- Straatsma T P and McCammon J A (1991), *J. Chem. Phys.* **95**, 1175-1188.
- Straatsma T P and McCammon J A (1992), *Ann. Rev. Phys. Chem.* **43**, 407-435.
- Tropsha A and Hermans J (1992), *Prot. Eng.* **5**, 29-33.
- van Gunsteren W F and Berendsen H J C (1987), *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos B.V., Nijenborgh 16, Groningen, The Netherlands.
- Veerapandian B *et al.* (1990), *J. Mol. Biol.* **216**, 1017-1029.
- Warshel A and Russell S T (1984), *Q. Rev. Biophys.* **17**, 283-422.
- Warshel A and Creighton S (1989), in: van Gunsteren, W.F. and Weiner, P.K. (eds.), *Computer Simulation of Biomolecular Systems*. ESCOM, Leiden, 120-138.
- Weiner S J *et al.* (1986), *J. Comput. Chem.* **7**, 230-252.
- Wong C F and McCammon J A (1986), *J. Am. Chem. Soc.* **108**, 3830-3832.

CLAIMS

1. A method for predicting the binding free energy between a host or receptor molecule and a chemical substance bound thereto, comprising

assessing the average energy difference, $\langle \Delta V_{i-s}^{el} \rangle$, defined as $\langle V_{i-s}^{el} \rangle_B - \langle V_{i-s}^{el} \rangle_B$, between the contribution from polar interactions to the potential energy between the chemical substance (denoted i) and its surroundings (denoted s) in two states, one state (A) being where the chemical substance is surrounded by solvent, the other state (B) being where the chemical substance, bound to the host or receptor molecule, is surrounded by solvent,

assessing the average energy difference, $\langle \Delta V_{i-s}^{vdw} \rangle$, defined as $\langle V_{i-s}^{vdw} \rangle_B - \langle V_{i-s}^{vdw} \rangle_A$, between the contribution from non-polar interactions to the potential energy between the chemical substance (denoted i) and its surroundings (denoted s) in two states, one state (A) being where the chemical substance is surrounded by solvent, the other state (B) being where the chemical substance, bound to the host or receptor molecule, is surrounded by solvent,

and calculating the absolute binding free energy as an adjusted combination of the two above-mentioned average energy differences.

2. A method according to claim 1, wherein the assessment of at least one of the average energy differences is performed by establishing a 3-dimensional representation comprising a 3-dimensional structure of the host or receptor molecule and a 3-dimensional structure of the chemical substance docked therein and applying molecular dynamics calculations to the 3-dimensional presentation.

3. A method according to claim 1, wherein the assessment of both of the average energy differences is performed by estab-

lishing a 3-dimensional representation comprising a 3-dimensional structure of the host or receptor molecule and a 3-dimensional structure of the chemical substance docked therein and applying molecular dynamics calculations to the 3-dimensional representation.

4. A method according to any of the preceding claims, wherein the free energy of binding is predicted as

$$\Delta G_{bind} = \beta<\Delta V^{el}_{i-s}> + \alpha<\Delta V^{vdw}_{i-s}> + c \qquad (2b)$$

where α and β are coefficients representing the result of a calibration established by comparing the results of predictions with actual measured values, and c is a constant reflecting extrapolation to zero size of the chemical substance or reflecting correction for possible systematic errors.

5. A method according to claim 4, wherein the free energy of binding is predicted as

$$\Delta G_{bind} = \beta<\Delta V^{el}_{i-s}> + \alpha<\Delta V^{vdw}_{i-s}> \qquad (1b)$$

6. A method according to claim 4 or 5, wherein the absolute value of α is at the most 1.0, preferably at the most 0.3.

7. A method according to claim 6, wherein the absolute value of α is at the most 0.2.

8. A method according to claim 7, wherein the absolute value of α is about 0.16.

9. A method according to claim 8, wherein the value of α is about 0.16.

10. A method according to any of claims 4-9, wherein the absolute value of β is at the most 1.0, preferably about 0.2 - 0.8.

11. A method according to claim 10, wherein the absolute value of β is about 0.3 - 0.7, preferably about 0.4 - 0.6.

12. A method according to claim 11, wherein the absolute value of β is about 0.5.

13. A method according to claim 12, wherein the value of β is about 0.5.

14. A method according to claim 13, wherein the free energy of binding is predicted as

$$\Delta G_{bind} = \tfrac{1}{2}<\Delta V_{i-s}^{el}> + \alpha<\Delta V_{i-s}^{vdw}> + c \qquad (2)$$

15. A method according to claim 14, wherein the free energy of binding is predicted as

$$\Delta G_{bind} = \tfrac{1}{2}<\Delta V_{i-s}^{el}> + \alpha<\Delta V_{i-s}^{vdw}> \qquad (1)$$

16. A method according to any of claims 4 or 6-14 wherein c is in the range -10 to 10 kcal/mol.

17. A method according to claim 16, wherein c is in the range of -3 to 3 kcal/mol.

18. A method according to claim 17, wherein c is in the range of -2 to 2 kcal/mol.

19. A method according to claim 18, wherein c is in the range of -1 to 1 kcal/mol.

20. A method according to any of the preceding claims, wherein the solvent is selected from organic aromatic and non-aromatic organic solvents, inorganic solvents and mixtures thereof.

21. A method according to claim 20, wherein the solvent is selected from methanol, ethanol, acetone, acetonitrile, chloroform, hexane, water and mixtures thereof.

22. A method according to any of claims 1-19, wherein the solvent is an aqueous solvent.

23. A method according to claim 22, wherein the solvent is water.

24. A method for identifying a chemical substance capable of interacting with a host or receptor molecule, e.g. binding to the host or receptor molecule, with a predicted binding energy equal to or better than a predetermined threshold value, comprising

1) choosing a chemical substance, A, which could potentially interact with the host or receptor molecule, and providing a 3-dimensional structural representation thereof,

2) predicting the binding free energy between the chemical substance A and the host or receptor molecule by the method according to any of claims 1-23,

3) if the predicted binding free energy between the chemical substance A and the host or receptor molecule determined is equal to or better than the predetermined threshold value, then identifying the chemical substance A as the chemical substance X,

4) if the predicted binding free energy between the chemical substance A and the host or receptor molecule determined is not equal to or better than a predetermined threshold value, modifying the 3-dimensional structural representation and predicting the binding free energy between the thus modified chemical substance, B, and the host or receptor molecule by the method according to any of claims 1-23, and

5) if necessary repeating step 4 until the predicted binding free energy determined between the resulting chemical substance, X, and the host or receptor molecule is equal to or better than the predetermined threshold value.

25. A method according to claim 24, further comprising providing a sample of a chemical substance or chemical substances selected from the chemical substances A, B and X and providing a sample of the host or receptor molecule, measuring the binding free energy between the chemical substance or substance and the host or receptor molecule, and if the measured binding free energy between the chemical substance or substances and the host or receptor molecule is not equal or substantially equal to the predicted value, then performing a calibration of the method according to any of claims 1-23 so as increase the predictive value of the method.

26. A method for providing a chemical substance capable of interacting with a host or receptor molecule, e.g. binding to the host or receptor molecule, with a binding energy equal to or better than a predetermined threshold value, comprising performing the method according to claim 23 or 24 to identify a chemical substance X having a predicted binding free energy equal to or better than the predetermined threshold value, providing a sample of the chemical substance X and a sample of the host or receptor molecule and measuring the binding free energy between the chemical substance X and the host or receptor molecule, and establishing that the measured binding free energy between the chemical substance X and the host or receptor molecule is equal to or better than the predetermined value.

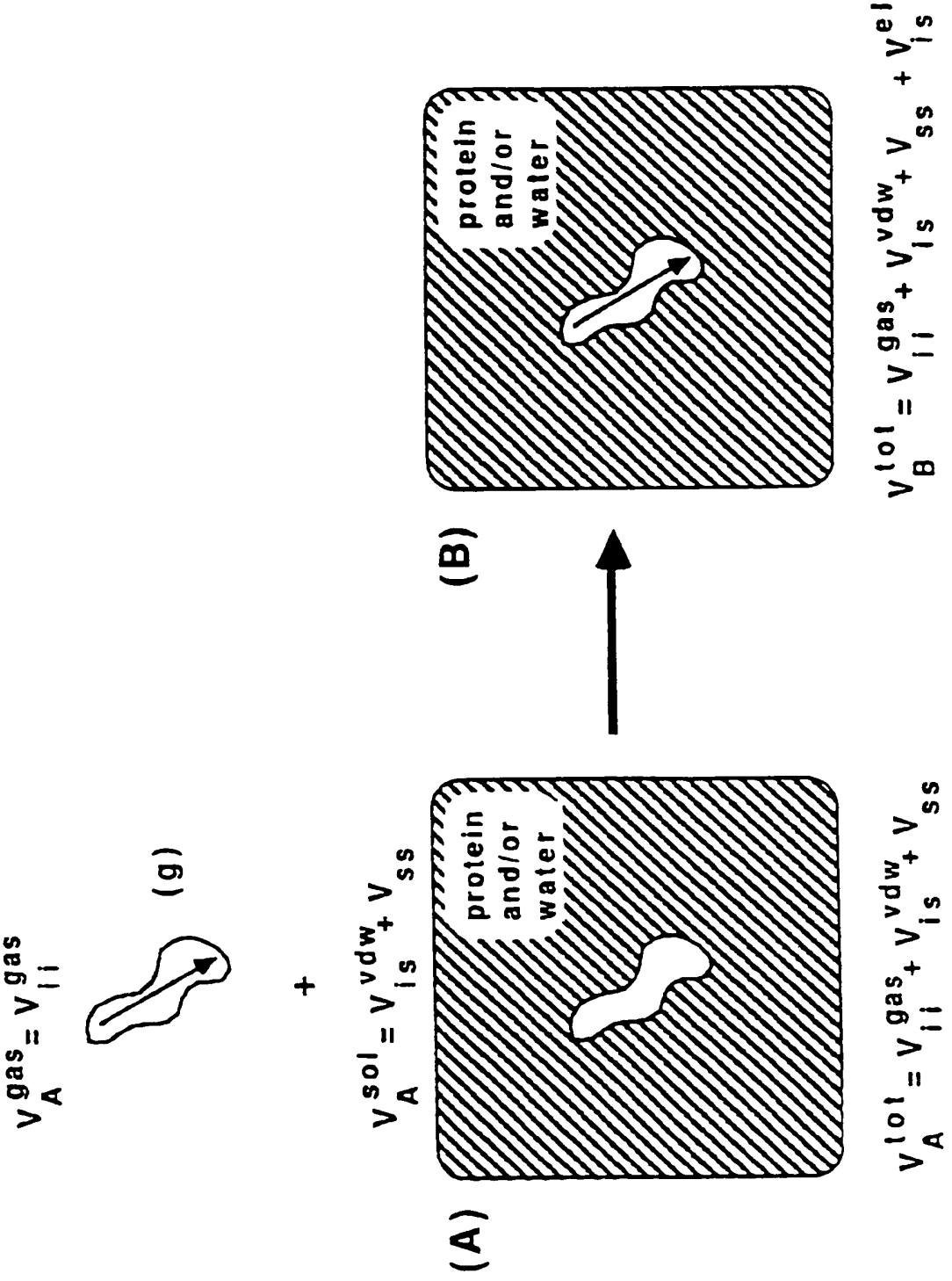

$V_A^{gas} = V_{ii}^{gas}$
(g)
+
$V_A^{sol} = V_{is}^{vdw} + V_{ss}$
protein and/or water
(A)
$V_A^{tot} = V_{ii}^{gas} + V_{is}^{vdw} + V_{ss}$
(B)
protein and/or water
$V_B^{tot} = V_{ii}^{gas} + V_{is}^{vdw} + V_{ss} + V_{is}^{el}$

Fig. 1

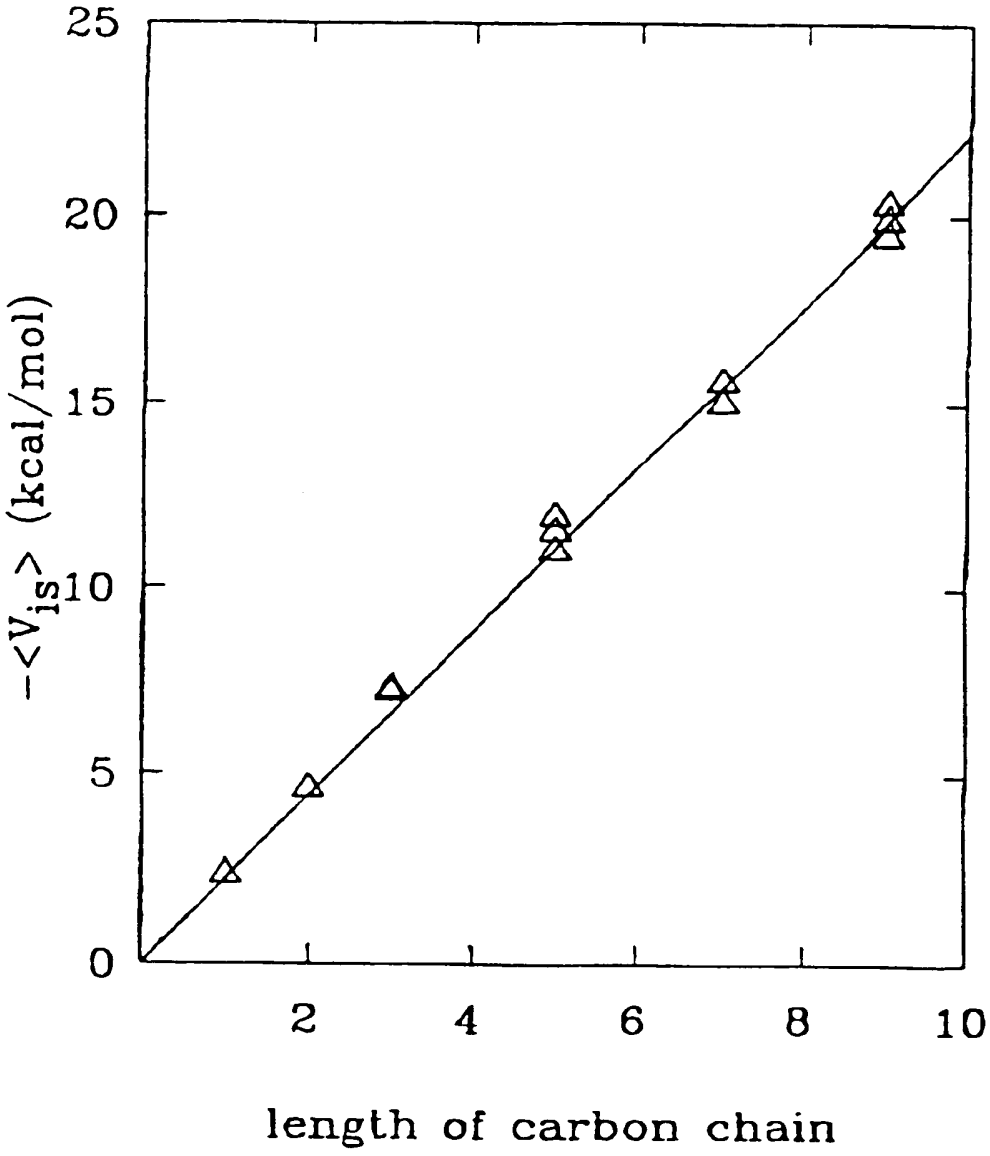
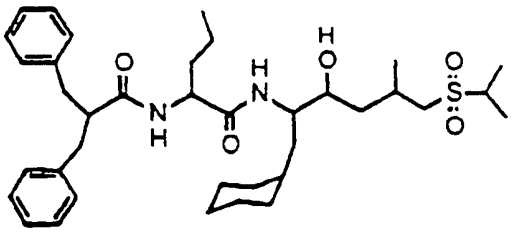
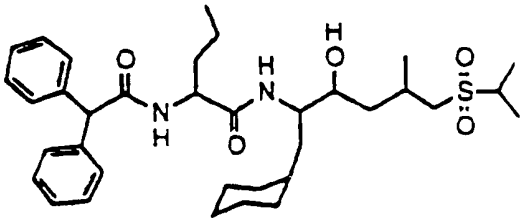
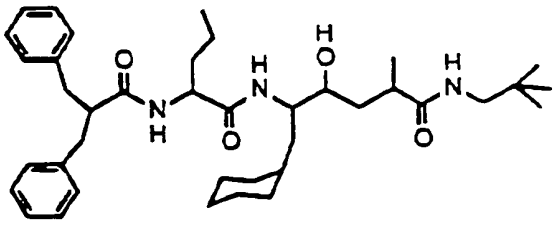
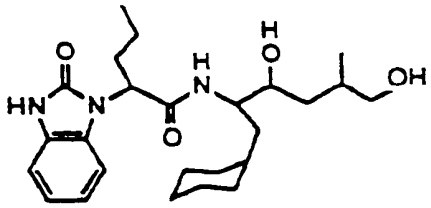


Fig. 2

I_1 

I_2 

I_3 

I_4 

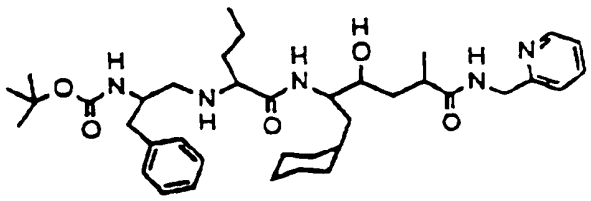
I_5 

Fig. 3

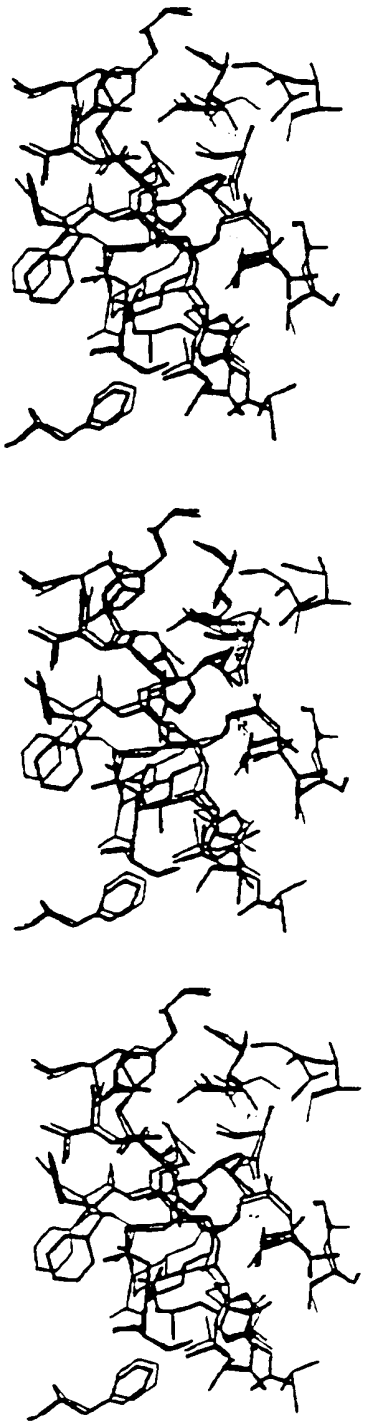


Fig. 4

INTERNATIONAL SEARCH REPORT

International Application No
PCT/IB 94/00257

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G06F17/50

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CHEMICAL ABSTRACTS, vol. 108, no. 1, 4 January 1988, Columbus, Ohio, US; abstract no. 2312, MILLER,S ET ALL. 'INTERIOR AND SURFACE OF MONOMERIC PROTEINS' page 226 ; see abstract & J.MOL.BIOL., vol.196, no.3, 1987, UK pages 641 - 656	1

☐ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 November 1994	15. 12. 94

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Stienon, P

Form PCT/ISA/210 (second sheet) (July 1992)